1. Suppose you want to store a petabyte of data, and you want to run a report that requires reading and processing 250 terabytes of that data. What is a key difference in the technology you'll use for this, versus a need to store and process one or two megabytes of data. **1 / 1 point**

- ○ Extremely powerful computer processors
- ○ Cost
- ○ Cloud storage
- ◉ Distributed storage and processing

✓ **Correct**

Correct. The solutions built today to store and process big data usually involve the use of many computer CPUs and many storage disks, perhaps counting into the thousands. This is true whether the data store and processing are handled locally or in the cloud. So, it is the nature of big data systems that the processing is necessarily handled not by one monolithic program that does all the work, but by a collection of computers that collectively do the data processing work.

2. The following are records in a contact list. **1 / 1 point**

{'name';'Étienne', 'email';'etienne@example.com', 'mobile';'555-8372'}

{'name';'Brayden', 'home';'555-2202', 'work';'555-2800'}

{'name';'Diana', 'mobile';'555-6575', 'email';'dprince@example.com'}

Is this contact list an example of structured, semi-structured, or unstructured data?

- ○ Structured
- ◉ Semi-structured
- ○ Unstructured

✓ **Correct**

Correct. These records don't always have the same information, and even when they do, the information isn't in the same order. There is some structure here, but not as much as we would expect from a structured table.

3. **1 / 1 point**

Is an online poll asking participants to rate a selection of statements from strongly disagree (1) to strongly agree (5) an example of a structured, semi-structured, or unstructured data?

- (●) Structured
- ( ) Semi-structured
- ( ) Unstructured

✓ **Correct**

Correct. Each record will simply be a sequence of numerical ratings corresponding to the sequence of statements.

4. You plan to gather data from various sources. Which of the following sources do you think will definitely give you structured data? **1 / 1 point**

- [x] A survey in which every question is a rating from 1 to 5

✓ **Correct**

Correct. This would most likely be structured data, with verified data in structured fields.

- [ ] A set of XML documents delivered from a public data source
- [ ] A news article downloaded from the web
- [ ] A CSV (comma separated value) file taken from a spreadsheet
- [x] Tables you capture from another relational database system

✓ **Correct**

Correct. A table in an RDBMS will be structured.

- [ ] A collection of photographs taken with a smartphone

5. Which of the following describe a reason why RDBMSs are a poor choice for big data? Check all that apply. **1 / 1 point**

- [ ] Because RDBMSs verify data on write, each row must be INSERTed separately, which is prohibitive when you have millions of rows

- [x] The structured nature of RDBMSs imposes costs in terms of storage and processing, which becomes prohibitive with really big amounts of data.

Correct

Correct. The cost per terabyte of data in an RDBMS can be 10 or even 100 times as much as the cost for data in a simple file store. Big data systems can work with and take advantage of these simple file storage systems.

- [x] Because a large amount of unstructured data would need to be stored as a BLOB or CLOB, RDBMSs provide little to no support for working with such data.

Correct

Correct. SQL provides very little means to search, sort, or calculate information from BLOB or CLOB data types, so large unstructured data would be virtually useless in an RDBMS system.

6. Look at the following data: **1 / 1 point**

id	name	grade_level	gpa	age
930	Olufunmilayo Ayton	11	4.00	16
667	Vincent Michaelson	10	2.53	15
907	Asa Quigg	10	3.57	
168	Kiran Patil	11	3.28	17

Now imagine that, instead of four rows, you have 4000 rows, and all are similar to the rows you see here. Which of the following questions can you answer from this data? Check all that apply.

- [x] What is the number of students in the table?

Correct

Correct. This would simply be a count of the number of rows.

- [x] What is the number of students in each grade level?

Correct

Correct. This could be done by finding how many rows have a particular value for grade level.

- [ ] What is the home address of a student with **id** '930'?
- [ ] What is the highest allowable value of **age**?
- [x] What are the names of all the students at the same grade level as Kiran Patil?

Correct

Correct. A query could first find the **grade_level** of 'Kiran Patil' and then find all other students with the same **grade_level** value.

7. Consider the following data (in this case, a list of JSON objects): **1 / 1 point**

{'shop':'Dicey', 'game':'Monopoly', 'qty':7, 'aisle':3, 'price':17.99}

{'shop':'Dicey', 'game':'Clue', 'qty':3, 'price':9.99}

{'shop':'Board Em', 'game':'Monopoly', 'qty':11, 'aisle':2, 'price':25.00}

{'shop':'Board Em', 'game':'Candy Land', 'qty':4, 'aisle':2}

{'shop':'Board Em', 'game':'Risk', 'qty':7, 'aisle':3, 'price':35.00}

{'shop':'Board Em', 'game':'Stratego', 'qty':'low stock'}

Which of the following questions can you definitely answer from this data? (Hint: Take note of missing values and inconsistent data types, which would make the answers unknown or uncertain.)

- [x] What are the games that start with the letter C?

Correct

Correct. You can use the game field to find games that start with a particular letter.

- [ ] Which shop has a higher average price for its games?
- [ ] Which games are in aisle 3 at the Dicey shop?
- [ ] What is the average quantity of all the games in all the stores?

- [x] What is the price of Risk at the Board Em shop?

Correct

Correct. This data is provided.

8. Which of the following questions could be answered quickly and easily by treating the complete plays of Shakespeare as data, separated by title and type (tragedy, comedy, or history)? Check two answers. **1 / 1 point**

- [x] Which of the tragedies includes, or mentions, someone named Lucilius?

Correct

Correct. Searching just the tragedies for 'Lucilius' will produce this information.

- [x] How many plays are histories?

Correct

Correct. Since they are separated by type, the actual contents of the plays do not need to be analyzed to answer this question.

- [ ] Which of the plays are considered the most important?
- [ ] How many people are mentioned or appear in the plays?